EXHIBIT 85

United States District Court

Northern District of California

Case No.: 3:18-cv-01586-JSC

IN RE PACIFIC FERTILITY CENTER LITIGATION

Expert Report of Nicholas P. Jewell, Ph.D.

October 15, 2019

Expert Report of Nicholas Jewell, Ph.D.

in Support of Plaintiffs' Motion for Class Certification

I. Introduction

1. For more than 38 years, I have been a Professor in the Division of Biostatistics, School of Public Health, and in the Department of Statistics, both at the University of California, Berkeley. Specifically, I have served as a Full Professor (1987 – 2018); Associate Professor (1983 – 1987); and Assistant Professor (1981 – 1983). I am currently Professor in the Graduate School at the University of California, Berkeley while recently assuming a position as Chair in Biostatistics and Epidemiology at the London School of Hygiene & Tropical Medicine in London, U.K.

2. Prior to coming to Berkeley, I was an Assistant Professor in the Department of Statistics at Princeton University, Princeton, New Jersey (1979 – 1981) where I also served as Director of the Statistical Laboratories.

3. At Berkeley, I also held the position of Chair of the University of California Graduate Group in Biostatistics from 1986 – 1994, and from 2000 – 2007. From 1994 to 2000, I served as Vice Provost at the University of California, Berkeley, in the Office of the Chancellor. From 2007 – 2008, I served as Vice Provost, Academic Personnel, at the Office of the President of the University of California. A recent copy of my CV is attached as Exhibit A.

4. I received my Ph.D. in Mathematics from the University of Edinburgh, Scotland in 1976, and was a post-doctoral Harkness Fellow at Stanford University and the University of California at Berkeley from 1976 to 1978 (funded by the Commonwealth Fund of New York). In 1978 – 1979, I was a Research Fellow in the Medical Statistics Unit at the University of Edinburgh, Scotland.

5. I have held Visiting Professor appointments at Oxford University, England, in Spring 1990, at the London School of Hygiene and Tropical Medicine in Spring 2007, at Kyoto University in Japan in Fall 2009, and at the Karolinska Institutet, Stockholm, in Fall 2016, 2017, and 2018. During April – May 2007, I was a resident Fellow of the Rockefeller Foundation at its Study Center in Bellagio, Italy.

6. I have also served as a member of the National Academy of Sciences Committee on National Statistics (1993 – 1996) and of the Committee on Theoretical and Applied Statistics (1994 – 1996).

7. I am the author of a best-selling textbook, "Statistics for Epidemiology" (Chapman and Hall, New York 2003), a recently co-authored book on causation, "Causal Inference in Statistics: A Primer" (John Wiley, 2016), as well as approximately 180 peer-reviewed articles in the field of biostatistics (see CV, attached as Exhibit A). My areas of expertise include the statistical design and analysis of studies used to investigate risk factors for disease outcomes including adverse effects using methods such as longitudinal and survival data analysis.

8. I recently stepped down from four years of service (2013 – 2017) as Editor for the Journal of the American Statistical Association, the preeminent statistical journal in the United States. I am also a founding editor of The International Journal of Biostatistics and Statistical Applications in Genetics and Molecular Biology. I am Associate Editor for the historic British journal Biometrika.

9. In 2005, I received the Snedecor Award, from the Committee of Presidents of the Statistical Societies, awarded to "an individual who was instrumental in the development of statistical theory in biometry." The award is associated with the best publication in biostatistics

in the world in the previous three years. In 2004, I received a Distinguished Teaching Award from the School of Public Health, University of California at Berkeley. I was the 2012 recipient of the Marvin Zelen Leadership Award in Statistical Science from Harvard University, and the 2018 recipient of the Adrienne Cupples Award for Excellence in Teaching, Research, and Service in Biostatistics from Boston University.

10. I was the 2018 Bradford Hill Lecturer at the London School of Hygiene & Tropical Medicine, and the 2019 Greenberg Lecturer at the University of North Carolina, Chapel Hill.

11. I was elected to the National Academy of Medicine in 2017. In addition, I am, or have been, a member and officer of several international statistical societies including the International Biometric Society, the Institute of Mathematical Statistics, and the American Statistical Association. I have served as Chair of the Section on Statistics in Epidemiology of the American Statistical Association (2009 – 2012). I was elected a Fellow of the American Statistical Association in 1991, a Fellow of the Institute of Mathematical Statistics in 1996, and am also a Fellow of the Royal Statistical Society in the UK. I served as President of the Western North American Region of the International Biometric Society in 1991 – 1992, and as Treasurer of the Institute of Mathematical Statistics from 1985 – 1988. In 2007, I was elected as Fellow of the American Association for the Advancement of Science ("AAAS"). In 2015, I served as Chair-Elect of the Statistics Section of the AAAS, as Chair in 2016, and as past-Chair in 2017. In 2015, I also became an elected member of the International Statistical Institute.

12. The opinions below are based on findings reached to a reasonable degree of scientific certainty. I reserve the right to amend this preliminary report to account for additional information.

13. My rate is $700 per hour including deposition and court testimony. Payment is not contingent upon the outcome of the case.

14. A history of my prior testimony in the past four years is attached as Exhibit B.

II. Background and Scope of Assignment

15. I understand that on March 4, 2018, laboratory staff at Pacific Fertility Center ("PFC") determined that a loss in liquid nitrogen levels in a cryopreservation tank (hereafter Tank 4) had occurred for an unknown amount of time, potentially damaging the eggs and embryos contained in the tank (the "Incident").

16. Based on available data, including information on the condition of eggs and embryos that have been subsequently thawed from Tank 4, I have been asked to describe and develop statistical methods to assess the general impact that the tank failure has to date had on the eggs and embryos that were exposed to this Incident. This necessarily requires a comparison of success rates of the eggs and embryos[1] in Tank 4 in comparison to analogous rates in the facility presumably for eggs and embryos not exposed to conditions experienced in Tank 4.

17. I have not been asked to provide opinions on the emotional or economic impact on class members, nor was I asked to assess the risks to children born from eggs and embryos that originated from Tank 4 after the incident.

18. It is my understanding that an embryologist retained by Plaintiffs, Christine Allen, will opine on concepts underlying in vitro fertilization (IVF) and, among other matters, the impact of the incident on reproductive tissue exposed to the incident. I further understand that Dr. Stephen Somkuti, an Obstetrician Gynecologist retained by Plaintiffs, will also provide testimony as to the nature of the IVF process and the IVF experience.

[1] I am using the term "success rates" as it is defined and set forth in the Expert Report of Christine Allen. It is my understanding that this is a term of art in the industry that refers to several benchmarks, including thaw success rates, clinical pregnancies, and live birth (or delivery) rates.

III. Summary of Preliminary Opinions

19. Using publicly available data extracted from the Society for Assisted Reproductive Technology (SART) website that is based on information supplied directly from Pacific Fertility Center, and analogous information regarding thaws of embryos from Tank 4 (following the Incident), I describe basic statistical methods to compare (i) the rate of producing a thawed embryo that is viable for transfer and implant, (ii) the rate of achieving a clinical pregnancy from a transferred embryo, and (iii) the rate of achieving a live birth from a clinical pregnancy.

20. My preliminary analyses directly compare these rates for the historical 2017 SART data to those for Tank 4 post-March 4, 2018, the latter data provided by defense counsel. This initial report discusses briefly some of the statistical considerations in extracting the data and performing valid comparisons.

21. I also provide a preliminary comparison of the overall rate of achieving a live birth per thaw procedure (using data produced to date), a figure that ultimately determines the number of frozen embryos needed, on average, to obtain a live birth.

22. My preliminary calculations show that (i) the rate of producing a thawed embryo that is viable for transfer is currently estimated to be *less than half* for Tank 4 embryos as compared to the 2017 data for PFC, (ii) the rate of achieving a clinical pregnancy from a transferred embryo from viable Tank 4 thawed embryos is also *about a half* of what is shown for 2017 PFC successful thaws, and (iii) the rate of achieving a live birth from a clinical pregnancy resulting from Tank 4 is again *about a half* of what is shown for 2017 PFC pregnancies.

23. Overall, the rate of live births per embryo thaw procedures for Tank 4 thaws is therefore *about one eighth* of what was historically shown from 2017 PFC clinic data. Specifically, this live birth rate is estimated to be 46.5% for the 2017 PFC data but only 5.8% for

the Tank 4 thaws, a statistically significantly lower rate. Based on my preliminary analysis, the observed different success rates are extremely unlikely to have arisen by chance.

24. I also express the impact of this sharp drop in the likelihood of achieving a live birth from a frozen Tank 4 embryo by indicating that estimates from the data suggest that, on average, 17.4 thaw procedures from Tank 4 are required to obtain a live birth, as compared to 1.9 thaws based on 2017 information. In other words, those who stored embryos in Tank 4 have lost *almost 90% of their frozen embryos* in terms of achieving the goal of a live birth based on expectations reflected in the reported 2017 data.

25. With regard to thaws alone, my calculations show that 94% of the failed thaws (i.e. thaws that did not lead to embryos viable for transfer) can be attributed to the Tank 4 incident, all other things being equal. In other words, 94% of the failed thaws would have been prevented, on average, had the successful thaw (i.e. thaws that led to an embryo viable for transfer) rate for Tank 4 embryos matched what was reported for the 2017 PFC thaws.

26. Further analyses will refine these calculations using additional information on characteristics of the embryos thawed both in 2017 and from Tank 4, particularly the woman's age at the time of egg extraction and the role of using a donor egg to create the embryo as compared to the woman's own egg. To date, I have not conducted analysis using age information, given the recent production and my ongoing analysis of that data. Future analyses will take this into account.

27. Further analyses will also compare success rates arising from frozen eggs (that are fertilized after thawing), although the available data on eggs is associated with fewer women than those receiving thawed embryos.

28. Using the methods and information described in this report, I can determine, to a reasonable degree of scientific certainty, the impact that the Incident had on the success rates for eggs and embryos in Tank 4. More specifically, I can determine the extent to which the Incident has reduced the likelihood of achieving a successful thaw, clinical pregnancy, and live birth. I can also estimate what percent of the failures that have occurred are statistically attributable to the Incident. Finally, I can also estimate the overall percentage of eggs and embryos that have been essentially "lost" as a result of the Incident, when viewed as how many eggs or embryos a person would have to thaw, on average, to achieve a live birth.

IV. Analysis

A. Measuring the Impact of the Incident on the Eggs and Embryos Stored in Tank 4.

1. Standard Statistical Methodologies for Comparing Outcomes Between Two Groups.

29. Often scientific interest focuses on a comparison of outcomes across two population groups. Here, the two groups might represent (i) embryos exposed to the Tank 4 incident and (ii) other embryos not so exposed in the same facility; for example, is the successful thaw probability the same in the two groups or is it lower for Tank 4 embryos? Standard statistical methodologies, described in detail below, can be applied to assess the differences in observed outcomes between these two groups, and to determine the extent to which differences among those groups can be attributed to exposure to the Incident, as opposed to other factors including chance.

30. A supplemental method to describe any difference in underlying probabilities of success is provided by the so-called *attributable risk* (among the exposed). I will also use this attributable risk methodology here. Refocusing on the probability of a "failure" (rather than success), the attributable risk describes the fraction of failures in one group (the "exposed"—

here, the Tank 4 eggs/embryos) that would have been prevented (i.e. been successes) if that group had enjoyed the (higher) success probability of the other group (the "unexposed", i.e. eggs/embryos not exposed to the Tank 4 incident).

31. For example, when comparing successful embryo thaw rates for two groups of embryos, suppose one observed a successful thaw rate of 85% in one group and 95% in another. For 100 embryos in the first group, this would reflect 15 unsuccessful thaws. One can then ask, what would be the fraction of those 15 failures that would have been successes had the true successful thaw rate been 95% as seen in the other group. This fraction is known as the *attributable risk among the exposed*.[2] So, in this example, if a success rate of 95% had been operating in the first group, only 5 unsuccessful thaws out of a 100 frozen embryos would have been expected (instead of the 15). Thus, two thirds—or 67% of the 15 failures in the first group can be attributed to this group not benefiting from the higher success rate of the second group. Since 67% > 50% in this example, more likely than not any failure in the first group can be attributable to the exposure conditions for that group. The attributable risk is a statistical measure and therefore applies to a statistical assessment of the entire egg/embryo population in the exposed group.

2. Determining Appropriate Outcome Measures for IVF Data

32. In all studies, it is necessary to determine appropriate outcome measures for comparison. It is my understanding that during the process of thawing a frozen embryo, successfully transferring it and implanting it in a woman, and then following the transfer though a potential pregnancy and desired live birth, there are several possible stages where "failures"

[2] See Jewell, *Statistics for Epidemiology*, Chapman & Hall/CRC, Ch. 9 (2004).

can occur, allowing a number of intermediate outcomes to assess impact.[3] Here, I specifically focus on three steps of the process where failure can occur. First, I consider the successful thaw outcome that describes whether or not a frozen embryo is successfully thawed and considered viable for implant in a woman—using available data, I thus focus on the estimate of the successful thaw rate for different embryo groups. Second, amongst successful thaws and transfers, I consider the probability of achieving a positive clinical pregnancy[4]—such a probability can also be compared across two or more embryo groups. Finally, amongst clinical pregnancies, I consider the probability of achieving a live birth from the pregnancy. This probability can also be compared across embryo groups.

33. Examining these distinct stages of the process from thawing to pregnancy to live birth does not preclude an overall comparison, say, of the probability that one thawed embryo ultimately leads to a live birth. This is because this probability is simply the product of the probabilities of the distinct stages: (i) the probability of the thawed embryo, for example, leading to a viable transfer to the woman, (ii) the probability that such a viable transfer leads to a clinical pregnancy, and (iii) the probability of a clinical pregnancy leading to a live birth. I intend therefore to consider an overall assessment of the entire process, supplemented by examining the effects of the Tank 4 incident on each of the three stages.

34. Similar comparative analyses can be implemented for the analogous process for frozen eggs. It is important to note at the outset that the number of women who thawed eggs that were exposed to the Tank 4 incident is considerably smaller than for frozen embryos. For

[3] I refer to the embryologist for a more detailed explanation of success rates and benchmarks in the cryopreservation and IVF process.
[4] I am using the term clinical pregnancy here as defined by Christine Allen. I understand this is a term of art in the IVF industry to refer to a pregnancy in which all four datapoints (positive hCG, intrauterine presence of gestational sac, sac with normal conceptus, and presence of a heartbeat) fall within normal ranges.

example, for Tank 4, there is data on 151 thawed eggs that arose from only 11 different women. Nevertheless, it is valuable to be comprehensive in reported comparisons.

3. Controlling for Bias and Precision.

35. In drawing inferences from data that arises from any sort of study, two statistical aspects of estimation must be considered when comparing outcomes across differing conditions: *bias* and *precision*. Both of these issues are key: bias refers to systematic ways in which data comparing outcomes in two settings may not reflect true differences, whereas precision refers to comparative data not reflecting the truth due to random (i.e. non-systematic) effects. Precision is largely controlled by ensuring that samples of individuals studied are sufficiently large to limit any impact of random variation. Expanding the sample size, however, cannot correct for bias in the study. A helpful analogy can be found in considering how a photograph may reflect evidence of an event. If the camera is not pointed in the correct direction, the viewing frame is not covering the relevant scene, or the lens suffers from considerable distortion, the resulting photograph may not accurately capture the scene—this is analogous to systematic bias. If, however, the image is not distorted in these ways, the reproduction may still suffer from lack of clarity due to random camera "shake" or poor focus. For these different reasons, the resulting photograph may not provide sufficient detail of the event of interest—this is analogous to imprecision. Thus, study *bias* reflects systematic distortion whereas *precision* reflects random perturbations in the summary data (i.e. focus). Methods to control these two challenges to provide an accurate photo are quite distinct, as are the methods to address the statistical impact of these issues associated with data collection and analysis.

36. There are several ways to control for bias. For example, in many drug trials, randomization is used to show evidence of the effect of a therapeutic regimen. Randomization forms the basis of a valid statistical comparison between different groups of patients assigned to

alternative therapies, in part, because it ensures balance between the groups on other key variables that may influence the outcome of interest: "Randomization tends to produce study groups comparable with respect to known and unknown risk factors, removes investigator bias in the allocation of participants, and guarantees that statistical tests will have valid significance levels."[5]

37. In many contexts, randomization is not logistically or ethically possible, e.g. in determining the effects of smoking on lung cancer. In such cases, observational data are used where the exposure is not assigned at random (e.g. smoking). An absence of balance across groups can arise in observational studies since comparative groups are determined by other factors, and not by randomization. A lack of balance on key predictive factors is known as confounding.[6]

38. In this assignment, the groups refer to eggs and embryos stored (i) in Tank 4 at PFC, and (ii) eggs and embryos stored at PFC that were not exposed to an incident such as occurred in Tank 4. The goal is to compare thawing/pregnancy outcomes for eggs and embryos thawed from Tank 4 (one group) to eggs and embryos thawed at PFC immediately prior to the March 2018 incident (the comparison group). By choosing the second group of thawed eggs and embryos close in time to March 2018, I control for potential confounding of this observational data by chronological time as earlier data may reflect different outcomes due to different procedures in play. I discuss the control of additional confounding effects as available data permits.

39. Statistical methods to compare treatments are designed specifically to account for background patient variation and thus address the challenge of *precision*. Shifts in an outcome

[5] Friedman *et al*., *Fundamentals of Clinical Trials*, 3rd Ed., Springer, Ch. 5 (1998), at 61. *Id.* at 47.
[6] Jewell, *Statistics for Epidemiology*, Chapman & Hall/CRC, Ch. 9 (2004).

variable are often captured from study data by comparing the estimated means of the outcome variables across comparative groups, allowing assessment of a difference. For example, the outcome for a comparison of weight-loss programs might be measured by weight-loss at the end of one year of treatment. The mean weight for individuals experiencing a particular treatment can be estimated from observed (varying) weight losses for a sample of individuals under that treatment; the analogous estimated mean for individuals experiencing a comparator treatment can similarly be calculated from another sample of such persons. The estimated *difference in means* is simply the difference in these two estimates. Of course, use of estimated means is not a perfect reflection of the true unknown impact of the difference in weight loss for the two treatments in a large population. This is because, as noted, there is random variation in observed weight-loss across individuals—often caused by unknown factors—even though they are treated identically. This random variation necessarily impacts the precision in any treatment comparison, e.g., by comparing means. One way statisticians are able to assess how far the estimate of the treatment difference is likely to be from the true difference is by calculating a *confidence interval* (CI) using knowledge of sampling properties and the impact of random variation in individual measurements on estimates of the relevant means. Properties of a CI are assessed through hypothetical repetitions of an identical study on different samples from the same patient population. The CI signals to a reader uncertainty that is due to patient-to-patient variation and from only observing a sample of individuals (rather than the entire population). An estimated difference of outcome means should thus be accompanied by its associated CI so that a reader understands the precision, or lack thereof, of the data.

40. The *confidence level* of a standard CI is usually 95%. An interpretation of a 95% CI is that the probability that this confidence interval contains the true unknown population value

(here, the true difference in mean weight loss across two populations) is 0.95. For example, when the estimated difference in weight loss associated with two dietary interventions is 10 lbs over a year with an associated 95% confidence interval of (-5 lbs, +25 lbs), based on a sample of 100 individuals (with 50 each exposed to each intervention), one can be 95% confident that the true average difference in weight loss (if applied to the entire population of interest) would lie somewhere in the range of -5 lbs to +25 lbs. Although reporting the data as a difference of means provides a single numerical summary, a wide confidence interval, as here, reflects that there is very little certainty that the estimated mean difference reflects the true value accurately, i.e. there is low precision. In using 95% confidence intervals, the statistician is tacitly accepting a 5% error rate. However, reducing this error rate (say by using a 99% CI) necessarily makes the confidence interval much wider and thereby less informative. Increasing the sample size is one method to reduce the width of a confidence interval since necessarily more data leads to more precision in our estimates (means are more precisely estimated from samples of 200 individuals than from samples of 100 individuals). Note that this interpretation of a confidence interval only deals with precision and thus assumes that there is no bias. In the presence of bias, an apparent 95% CI may contain the true value with probability much smaller than 0.95.

41. An alternative measure of the strength of evidence against a straw *null hypothesis* is given by a *p-value*. The null hypothesis is usually that there is no difference in the distribution of outcome values in different comparative groups, specifically as captured by the means of the outcome distributions. If a null hypothesis is true, the population means would not differ and any estimate of the difference of means should be close to zero, reflecting the underlying true value. Of course, results based on a small sample of subjects, will not show an exact difference of zero but will be impacted by random patient variation as discussed earlier. Statisticians use the

estimate and their calculation of this anticipated variation to assess whether the belief that the true difference is zero is reasonable given the data that has been observed; this assessment yields a p- value that captures the strength of evidence against the assumption that the true difference is zero. One interpretation of the p-value is thus the probability of observing as large (or larger) a difference of estimated means as seen in the data *given the null hypothesis of no population difference in the means is true.* (It may help to think of a p-value as a measure of the probability that a suspect's fingerprints are on the weapon (or even worse evidence), *given that the suspect did not commit the crime.)* A small p-value thus reflects a small possibility that there is no true difference between the population means. The logic of statistical inference is thus to take small p-values as an indication that our assumption of no difference in population means is probably incorrect. (In our weapon/crime analogy, a small probability regarding the presence of the suspect's fingerprints leads to the conclusion that they likely committed the crime, i.e., rejection of the null hypothesis of innocence.) It is conventional in single efficacy trials to use 0.05 as a threshold for p-values for accepting the data as statistically significant, i.e., leading to a rejection of the null hypothesis (or not). As noted above for CIs, the relevance of the p-value is again predicated on an assumption that there is *no bias* in using the sample data to estimate population means, for example. Further, as for a 95% CI, accepting a p-value threshold of 0.05 means that an investigator has accepted a 5% false positive rate in declaring a *statistically significant* difference (i.e. p-value < 0.05) as evidence of a true difference caused by the therapy.

42. With small amounts of data, it is necessarily difficult to achieve a statistically significant difference on outcome patterns across comparative groups for a specified difference due to the lack of precision. In such cases, the absence of a small p-value reflects a lack of information rather than necessarily the absence of an important difference between groups. It is

important to note that as the size of a study sample increases, the amount of information about the effect of comparative therapies also increases and so the precision of a comparison between treatments improves.

43. A particular kind of outcome variable uses a simple binary indicator to measure a simple outcome, e.g. "success" or "failure".[7] In this case, the mean of the outcome is simply the probability of "success". For example, a binary outcome might be whether a frozen embryo was thawed successfully and was viable for transfer to a woman, or whether a viable transfer resulted in a clinical pregnancy, or whether a clinical pregnancy resulted in a live birth. In these cases, for a specified population or group of embryos, the average value of an outcome variable that measures "success"[8] is simply the probability of "success."

44. With either a continuous or binary outcome, statistical principles[9] allow us to derive estimates of the unknown population mean or probability, respectively, based on well-defined samples of individuals from the population of interest under specified assumptions. In addition, these methods also allow a measure of uncertainty of our estimate. The estimate, together with the uncertainty estimate, allow calculation of p-values for given null hypotheses (e.g. that the probability is 0.5 when testing whether a coin is fair) based on the observed data and the construction of confidence intervals. I will use this standard methodology, as discussed in more detail below, throughout.

[7] The logic of statistical inference remains the same with other kinds of outcome variables although the computational and statistical details necessarily change.
[8] With "success" coded as the value "1," and failure as the value "0".
[9] For example, the methodology known as *maximum likelihood*—see Jewell, *Statistics for Epidemiology*, Chapman & Hall/CRC, Ch. 13 (2004).

4. Data Sources for Measuring Impact of the Incident on Tank 4 Eggs and Embryos

45. The primary embryo group of interest is the set of frozen embryos stored in Tank 4 at PFC and subsequently thawed since the tank failure on March 4, 2018. Throughout this preliminary analysis, I assume that Tank 4 embryos that have been thawed since the Incident are representative of all Tank 4 stored embryos, absent any evidence to the contrary.[10] For comparative reasons, other groups of embryos refer to those from the facility that were thawed in 2017 and 2016, the most recent years before the tank failure.

46. I will consider two data sources regarding the outcomes associated with thawing and implanting embryos to achieve a successful pregnancy outcome.[11] The first is extracted from national In Vitro Fertilization (IVF) clinic information that is collated and provided by the Society for Assisted Reproductive Technology (SART). Specifically, I will focus on the preliminary data for Reporting Years 2016 and 2017, although all comparisons I make could also use earlier data when and if appropriate. Note that the SART labels the 2017 data as Preliminary and the 2016 data as Final. It is my understanding that the 2017 data is preliminary to the extent that the process for some women may not be complete with regard to determining all reported outcomes (given that a pregnancy lasts 9 months), and the final data will be reported following that period. My analyses in later reports will use Final 2017 data, 2016 data, and any preliminary 2018 data if and when it becomes available.

47. The second source of data was provided to counsel by Defendants and purportedly states the results of all thaw attempts from Tank 4 since March 4, 2018.[12]

---

[10] To date, I have not received any information on characteristics of *all* embryos stored in Tank 4 at the time of the Incident.

[11] I understand the Centers for Disease Control (CDC) also presents the equivalent data on clinic IVF outcomes in a different format that is thus duplicative of the SART data. I will consider exploration of the CDC information when the latter is produced.

[12] MSO025597-25600.

Additional data including results for aneuploid embryos (that are chromosomally abnormal) was recently produced and is not included in my analysis in this preliminary report. These results can be included in future calculations. I note that aneuploid embryo data would only affect the thawing success rates, as I understand these embryos are not viable for transfer.

48. Note that both the SART and Tank 4 data report thawing and pregnancy test outcomes by cycle, rather than by patient. To my understanding, current IVF guidelines recommend transferring a single embryo and that appears to be the most common scenario, but in some cases a physician may order two or more embryos to be transferred. To achieve a successful single embryo transfer, a clinic may have to thaw more than one embryo if the first embryo thawed is not viable for transfer. In that case, additional embryos are thawed until one successful thaw is obtained. If two embryos are requested, the process is continued until a second successful thaw that is viable for transfer is obtained.

49. SART sub-classifies data by age and whether eggs/embryos derive from a third-party "donor." This allows consideration of these factors in making comparisons as warranted. I only recently received age information for the Tank 4 data, and have not analyzed that data for purposes of this preliminary report. Future analyses will adjust comparisons for any differences in age distribution across the two groups. The Tank 4 data provided to date also does not distinguish embryos by egg source, so that it may contain mixed data that includes both a patient's own eggs and embryos fertilized using a donor's egg. At this point, I do not consider that the inclusion of embryos using donor eggs in the Tank 4 data would materially change my preliminary findings—this is because it is my understanding that donor eggs tend to come from younger women who have undergone more screenings so that better pregnancy results would be more likely all other things being equal. With this assumption, my conclusions are conservative,

in that the Tank 4 data should provide better success rates than if the donor egg embryos were excluded in comparison with the SART data.

50. Finally, there is data in both sources on frozen eggs in addition to frozen embryos, although for my preliminary analyses I focus on frozen embryos.

B. Preliminary Findings Demonstrate a Statistically Significant Drop in Success Rates for Embryos in Tank 4.

51. For a preliminary analysis and illustration of the application of statistical techniques to the data, I focus on SART data for outcomes associated with thawing frozen embryos using a patient's own eggs. For the preliminary data for 2017, Table 1 below shows the number of thaw procedures by the patient's age, along with information on the number of such procedures that did not yield thawed embryos that were deemed viable for transfer. Note that the SART summary data does not show how many of the cycle starts--where it was intended to implant an embryo--correspond to the *same* patient. We can thus only compute percentages etc. across embryos.

52. The preliminary data for PFC from 2017 is shown in Table 1:

	Age of woman (years)					
	< 35	**35--37**	**38--40**	**41--42**	**>42**	**Total**
Number of thaw procedures	86	89	107	54	34	370
Thaw procedures yielding embryos suitable for transfer	84	87	103	53	31	358
% of thaws yielding a viable transfer	97.7%	97.8%	96.3%	98.1%	91.2%	96.8%
Viable thaw procedures yielding a positive pregnancy test	59*	62*	62*	36*	26*	245
% of viable transfers yielding a positive pregnancy test	70.2%	71.3%	60.2%	67.9%	83.9%	68.4%
Viable thaw procedures yielding a clinical pregnancy	48*	53*	57*	28*	26*	212
% of viable transfers yielding a clinical pregnancy	57.1%	60.9%	55.3%	52.8%	83.9%	59.2%
Pregnancies yielding at least one live birth	40	43	43	21	19	166
% of clinical pregnancies yielding at least one live birth	83.3%	81.1%	75.4%	75.0%	73.1%	78.3%
# of live births	42	45	44	22	19	172
% of live births per thaw procedure	48.8%	50.6%	41.1%	40.7%	55.9%	46.5%

Table 1. 2017 SART data: patient's own eggs (frozen cycles). Extracted from online SART summary report for Pacific Fertility Center.
*** SART reports the positive pregnancy test and clinical pregnancy percentage based on the total number of thaw procedures.**

	Age of woman (years)					
	< 35	**35--37**	**38--40**	**41--42**	**>42**	**Total**
Number of thaw procedures	11	35	37	23	85	191
Thaw procedures yielding embryos suitable for transfer	5	16	16	11	36	84
% of thaws yielding a viable transfer	45.5%	45.7%	43.2%	47.8%	42.4%	44.0%
Viable thaw procedures yielding a positive pregnancy test	5	8	4	4	16	37
% of viable transfers yielding a positive pregnancy test	100%	50.0%	25.0%	36.4%	44.4%	44.0%
Viable thaw procedures yielding a clinical pregnancy	4	2	3	4	13	26
% of viable transfers yielding a clinical pregnancy	80.0%	12.5%	18.8%	36.4%	36.1%	31.0%
Pregnancies yielding at least one live birth	1	1	1	1	7	11
% of clinical pregnancies yielding at least one live birth	25.0%	50.0%	33.3%	25.0%	53.8%	78.3%
# of live births	1	1	1	1	7	11
% of live births per thaw procedure	9.1%	2.9%	2.7%	4.3%	8.2%	5.8%

Table 2. Tank 4 data on thaws since the Incident.

53. Table 2 shows the cumulative number of Tank 4 thaw procedures including how many of them were considered viable for transfer, again stratified by age.

54. With such data at hand, and subject to the limitations previously mentioned, we can now, in principle, compare success rates for three key stages of the process: (i) successful

thaws that yield a viable embryo for transfer, (ii) a successful clinical pregnancy after a viable transfer, and (iii) a live birth following a clinical pregnancy. We give some preliminary comparisons in this direction, although a full analysis awaits more complete data, including potentially the raw data underlying the SART figures, and clarification of existing data from PFC. While the conclusions are preliminary, the methods described in this report can be applied reliably to account for additional data, including additional age information and donor status.

55. The statistical analyses below represent standard statistical analyses of proportions as covered in early chapters of my textbook.[13] Certain independence assumptions are assumed as noted and additional statistical details can be found in the Appendix 1 that I have provided.

1. Viable Transfer Rate from Thaws

56. From Table 1, the estimated successful thaw rate from the 2017 PFC data is 97% with an exact 95% confidence interval of (94%—98%).[14] However, the analogous estimated successful thaw rate to date from Tank 4 embryos is 44% with a 95% confidence interval of (37%—51%).[15] Comparing the two underlying rates statistically with regard to their potential equivalence yields a p-value < 0.00001. Thus, the observed difference in thaw rates is extremely unlikely to have arisen by chance in a situation where the true underlying thaw success rates are identical. In other words, there is already striking evidence that the successful embryo thaw rate

[13] Jewell, *Statistics for Epidemiology*, Chapman & Hall/CRC, (2004).

[14] See Appendix 1 where the negative binomial distribution for the number of thaws used to obtain 1 or 2 viable transfers is considered. The calculations in this section assume that thaw procedures to obtain one or two viable transfers are independent whereas for in some cases, the process was repeated for a woman, presumably because the first transfers were not successful in yielding a live birth. For example, for the Tank 4 data, 20 of the 85 women "returned" for additional thaws. This will slightly decrease precision to the extent there is any correlation for thaw procedures that are repeated on the same woman. However, at this stage, it is not possible to adjust for this effect for the 2017 SART summary data that does not provide this level of detail. Future analyses will adjust for this effect when appropriate levels of detailed data are provided. Based on the Tank 4 data, this effect will not materially change my findings.

[15] See Appendix 1 for statistical details for these and other statistical calculations.

is much lower for Tank 4 embryos than what was apparent for the report of 2017 PFC thaws according to the SART data source.

57. With regard to the effect of age on the large difference in successful thaw rates, future detailed analyses can adjust for the specific effects of age using a more a complex form of analysis.[16] Such an analysis will include clear definitions of the age variables in both data sources to ensure comparability.

58. The attributable risk among the exposed (i.e. frozen embryos in Tank 4) is estimated to be 94% with an exact 95% confidence interval of (90%--97%). That is, on the surface, we can attribute 94% of the failed thaws to the Tank 4 incident: in other words, 94% of the failed thaws would have been prevented, on average, had the successful thaw rate for Tank 4 embryos matched what was reported for the report 2017 PFC thaws.

59. Given the estimated successful thaw rates of 97% and 44% for PFC thaws in 2017 and the Tank 4 thaws, respectively, another way of expressing the impact of this very large attributable risk is that according to the 2017 SART data, a woman would, on average, need to thaw 1.03 embryos to obtain 1 embryo viable for transfer; on the other hand, for Tank 4 embryos, the corresponding figure requires 2.3 embryos to be thawed, on average, to achieve 1 embryo viable for transfer. Effectively, women who stored embryos in Tank 4 have lost more than half of these embryos solely with regard to embryos viable for transfer. I now consider a preliminary assessment of the additional impact of the Tank 4 incident on obtaining clinical pregnancies and live births.

2. Clinical Pregnancy Test Rate from Viable Thaws

60. From Table 1, the estimated cumulative clinical pregnancy probability amongst successful thaws is 212/358 = 59.2% with an exact 95% confidence interval of (53.9%—

[16] Jewell, *Statistics for Epidemiology*, Chapman & Hall/CRC, Chapters 9—14, (2004).

64.4%)[17]. For the analogous Tank 4 data from Table 2, the estimated cumulative positive clinical pregnancy probability *amongst successful thaws*[18] is 26/84 = 31.0% with an exact 95% confidence interval of (21.3%—42.0%).[19] These can be compared statistically to determine the evidence that the clinical pregnancy rate is lower from Tank 4 successful thaws. This comparison again yields an exact p-value < 0.0001 indicating that the almost two-fold increase in successful clinical pregnancy rates (in 2017 versus Tank 4 successful thaws) is highly unlikely to have arisen by chance if the clinical pregnancy rates were identical in the two groups.

61. With regard to the effect of age on the difference in successful clinical pregnancy, future detailed analyses can adjust for the specific effects of age using a more a complex form of analysis. Such an analysis will include clear definitions of the age variables in both data sources to ensure comparability.

3. Live Birth Rate from Clinical Pregnancies

62. From Table 1, the SART data for PFC shows a live birth rate (of least one baby) of 166/212 = 78.3% per person, with an exact 95% confidence interval of (72.1%—83.7%). For Tank 4 data, the data produced to date indicates there were 11 live births from 26 clinical pregnancies, yielding a live birth rate of 11/26 = 42.3% with an exact 95% confidence interval of (23.4%—63.1%).[20] These similar rates can be compared statistically and show that the decline in

[17] This slightly understates the clinical pregnancy rate per woman if any women received more than one embryo simultaneously.
[18] The number of women who received these successful transfers was 77 since 7 women had two thawed embryos viable for transfer. Thus, the successful pregnancy rate *per woman* was 33.8%.
[19] The calculations in this section again assume that transfers are independent whereas, as previously, noted some women had more than one attempt at achieving a pregnancy. For Tank 4 data, 15 of 78 women (with viable transfers) had repeated attempts at achieving a clinical pregnancy. As noted earlier, at this stage, it is not possible to adjust for any correlation within women in results for the 2017 SART summary data as the summary source does not provide this level of detail. Future analyses will adjust for this effect when appropriate levels of detailed data are provided. Based on the Tank 4 data, this effect will not materially change my findings.
[20] The calculations in this section again assume that clinical pregnancies are independent. In this case, none of the 26 pregnancies for the Tank 4 data occurred in the same woman, suggesting that an independence assumption is appropriate here.

live births from clinical pregnancies associated with Tank 4 is statistically significantly different from what was reported in 2017 (p-value = 0.0002).

63. The attributable risk among the exposed (i.e. frozen embryos in Tank 4) is estimated to be 62% with an exact 95% confidence interval of (43%—75%). That is, on the surface, we can attribute 62% of the failed clinical pregnancies to the Tank 4 incident: in other words, an estimated 62% of the 15 failed clinical pregnancies associated with Tank 4 would have been prevented, on average, had the successful thaw rate for Tank 4 embryos matched what was reported for the 2017 PFC thaws.

64. With regard to the effect of age on the difference in live birth rates per clinical pregnancy, future detailed analyses can adjust for the specific effects of age using a more a complex form of analysis. Such an analysis will include clear definitions of the age variables in both data sources to ensure comparability.

4. Overall Comparison

65. For an overall comparison, we finally consider the total number of live births that resulted from thaw procedures, recognizing that a particular woman may have experienced multiple live births (likely from more than one implanted embryo). From Table 1, for 2017, 172 live births resulted from 370 embryo thaws, an overall live birth rate per thaw of 46.5% with an exact 95% confidence interval of (41.3%—51.7%). However, for Tank 4 embryo thaws, the data produced to date includes only 11 live births achieved from 191 thaws, a rate of 5.8% with an exact 95% confidence interval of (2.9%—10.1%).[21] Comparing these two overall rates of a live birth pregnancy per thaw yields a p-value < 0.0001.[22]

[21] The calculations for live birth rate per clinical pregnancy and thaw are based on the data produced to date, and will be adjusted to account for forthcoming data as outcomes are reported.

[22] The calculations in this section assume that thaw procedures to obtain live births are independent whereas for in some cases, the process was repeated for a woman as noted earlier. However, at this stage, it is not possible to adjust

66. Using similar calculations as described above, with estimated live birth rates/thaw of 46.5% and 5.8% for the PFC thaws in 2017 and the Tank 4 thaws, respectively, according to the 2017 SART data, a woman would, on average, need to thaw 2.2 embryos to obtain 1 live birth; on the other hand, for Tank 4 embryos, the corresponding figure requires 17.4 embryos to be thawed, on average, to achieve one live birth pregnancy. Effectively, women who stored embryos in Tank 4 have lost *almost 90% of their frozen embryos* in terms of achieving the goal of a live birth (88% to be more exact).

67. With regard to the effect of age on the difference in live birth rates per thaw procedure, there is no systematic effect of age in the 2017 SART data, future detailed analyses can adjust for the specific effects of age using a more a complex form of analysis. Such an analysis will include clear definitions of the age variables in both data sources to ensure comparability.

Dated: October 15, 2019

Nicholas P. Jewell Ph.D.

for this effect for the 2017 SART summary data that does not provide this level of detail. Future analyses will adjust for this effect when appropriate levels of detailed data are provided. Based on the Tank 4 data, this effect will not materially change my findings.

Appendix I: Statistical Details Underlying Group Comparisons

When estimating a population proportion of a "success," denoted say by *p*, the standard assumption is that a sample of size *n* is obtained from the population by simple random sampling. In this straightforward case, the maximum likelihood estimate of *p* is simply the observed proportion of success in the sample of *n* patients, that is n_1/n where n_1 is the number of patients in the sample who experience success. The standard error of this estimate is also straightforward and can be estimated by $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$].

For example, the cumulative number of 2017 cycle thaw procedures from the Pacific Fertility Center is 370 (see Table 1). Of these, 12 did not produce an embryo suitable for transfer (also see Table 1) so that the estimated successful thaw rate is 358/370 = 0.97, or 97%, with an apparent estimated standard error of 0.009. This yields a 95% confidence interval of (0.94—0.98) for the successful thaw fraction.

For the analogous Tank 4 data, the estimated successful thaw rate is, however, 84/191 (see Table 2) which is 0.44, or 44%, with an apparent estimated standard error of 0.036, and a 95% confidence interval of (0.37—0.51).

It is immediately obvious that these two estimated successful thaw fractions differ by far more than what would be expected by chance. This can be seen by the fact that the two 95% confidence intervals are distant from each other. More, directly, the underlying successful thaw rates can be compared statistically, yielding a p-value < 0.0001, reflecting that the observed difference in thaw rates is extremely unlikely to have arisen by chance in a situation where the true underlying thaw rates are identical.

The attributable risk among the exposed (i.e. frozen embryos in Tank 4) is estimated to be 94% with an exact 95% confidence interval of (90%--97%). That is, on the surface, we can attribute 94% of the failed thaws to Tank 4 issues; in other words, 94% of the failed thaws would have been prevented, on average, had the successful thaw rate for Tank 4 embryos matched what was reported for 2017 PFC thaws.

In the above analysis, I indicated the apparent standard error as, of course, it is clear that the data does not exactly arise from a simple random sample of *n* embryos where each embryo cycle is independent from the others (in particular, multiple embryos may be thawed for the *same* woman) and either leads to a successful transfer or not. In fact, as discussed above, the data from Tank 4, suggests that embryos are thawed, and then considered whether it is viable for transfer, before implant. For a particular woman, this transferred may be repeated several times until one or more viable transfers is obtained as per request from the IVF physician. For example, in Tank 4 for patient id #283492, 5 embryos were thawed before one was determined viable for transfer.

Some of the data can then be considered as the number of cycle thaws need to produce one viable transfer. The distribution of the cycle counts in this case follows what is known as the *geometric distribution*, a special case of the *negative binomial distribution*.

For other patients, it is clear that the process stopped before any successful transfer was obtained. See, for example, patient id 486719[23] where 6 unsuccessful thaws were carried out before the process stopped. Such data contributes what is known as a censored observation of the geometric distribution, that is we only know that more than 6 thaws would be required until one successful transfer was obtained, but not the exact number.

In addition, in a few Tank 4 case, more than one successful transfer was obtained. For example, patient id 491266[24], 2 thaws were initiated with 2 successful thaws. While the details of which embryos were transferred in which cycle are not provided with the limited data given, I have assumed that, for such patients, the data shows the number of transfers observed before 2 successful transfers were obtained. This random thaw count then follows a negative binomial distribution.

The three kinds of data, i.e. thaw counts until 0, 1, or 2 successful thaws were obtained, can be combined to yield a summary estimate of the successful thaw rate per cycle. It is straightforward to show that the maximum likelihood estimator for the successful thaw rate per cycle, *p*, is estimated by:

$$\hat{p} = \frac{n_1 + 2n_2}{T_0 + T_1 + T_2} \qquad (1)$$

where n_0 is the number of women who ultimately received no viable transfers, $n_1$is the number of women for whom one viable transfer was obtained, n_2 is the number of women for whom two viable transfers were obtained, T_0 is the total number of thaws associated with the n_0 women, T_1 is the total number of thaws associated with the n_1 women, and T_2 is the total number of thaws associated with the n_2 women.

It is immediately apparent that the numerator of this $\hat{p}$ is the total number of viable transfers from the thaws, so that $\hat{p}$ is simply the naïve fraction of all thaws that yield viable transfers, considered above. In the example, $\hat{p} = 0.44$ for the Tank 4 data.

[23] Identified as such in MSO_022465; however, this woman was apparently identified as patient id 155006 in MSO025597-25600.
[24] Identified as such in MSO_022465; this woman was apparently identified as patient id 156454 in MSO025597-25600.

The uncertainty of this estimator (1) can be obtained from what is known as the observed information that is derived from the likelihood function.[25] This yields the following formula for the standard error of the estimator $\hat{p}$ from (1):

$$\widehat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}^2(1-\hat{p})^2}{A(1-\hat{p})^2 + B\hat{p}^2}} \qquad (2)$$

where $A = n_1 + 2n_2$ and $B = (T_0 + T_1 + T_2) - (n_1 + 2n_2)$.

For the Tank 4 data, $n_1 = 70, n_2 = 7, T_0 = 53, T_1 = 113$, and $T_2 = 25$. Thus, as noted, $\hat{p} = \frac{84}{191} = 0.44$—from (1)--as with the naïve method. The standard error estimate from (2) = 0.036, essentially indistinguishable from the naïve standard error noted above.[26]

In addition, the standard error (2) assumes that thaw procedures to obtain one or two viable transfers are independent whereas for in some cases, the process was repeated for a woman, presumably because the first transfers were not successful in yielding a live birth. For example, for the Tank 4 data, 20 of the 85 women "returned" for additional thaws. This will slightly decrease precision to the extent there is any correlation for thaw procedures that are repeated on the same woman. However, at this stage, it is not possible to adjust for this effect for the 2017 SART summary data that does not provide this level of detail. Future analyses will adjust for this effect when appropriate levels of detailed data are provided. Based on the Tank 4 data, this effect will not materially change my findings.

Finally, none of the reported statistical estimates and comparisons adjust for age. Future detailed analyses can adjust for the specific effects of age using a more a complex form of analysis. Such an analysis will include clear definitions of the age variables in both data sources to ensure comparability.

[25] Observed information should be used here rather than expected information since the uncertainty estimate should not depend on properties of the variable that determines when transfers were stopped for women who ultimately received no viable transfers. In addition, this information is unknown, even for Tank 4 data.

[26] The naïve standard error estimate is 0.0359155 whereas the more complex estimate (2) yields 0.0359191, inconsequentially larger.

Appendix II: Scholarly Works Cited

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). Chapter 5: The Randomization Process. In *Fundamentals of Clinical Trials* (pp. 61-79). New York: Springer-Verlag.

Jewell, N. P. (2004). *Statistics for Epidemiology*. New York: Chapman & Hall /CRC.

Jewell, N. P. (2004). Chapter 5: Study Designs. In *Statistics for Epidemiology* (pp. 43-56). New York: Chapman & Hall /CRC.

Jewell, N. P. (2004). Chapter 9: Control of Extraneous Factors. In *Statistics for Epidemiology* (pp. 123-144). New York: Chapman & Hall /CRC.

Jewell, N. P. (2004). Chapter 13: Control of Estimation of Logistical Regression Model Parameters. In *Statistics for Epidemiology* (pp. 199-219). New York: Chapman & Hall /CRC.

Jewell, N. P. (2004). Chapter 9-14. In *Statistics for Epidemiology* (pp. 123-240). New York: Chapman & Hall /CRC.

Appendix III: Materials Relied On

I relied on the following in preparing the above report:

Publicly available data:

Society for Assisted Reproductive Technology (SART) website; data extracted from online summary report for Pacific Fertility Center available at https://sart.org.

Documents (by beginning bates):

MSO022465;

MSO025597